

## Trends in U.S. Foreign Policy Prioritizations, 2000 – 2019

Frederic Lestina  
American University  
[fl3300a@american.edu](mailto:fl3300a@american.edu)

### Abstract

*This paper examines trends in U.S. foreign policy priorities by recent U.S. presidents, using transcripts from Congressional foreign appropriations committees from 2000 to 2019. Textual analysis of the transcripts shows a divergence in distribution of key phrases, suggesting a possible shift in foreign policy focus by president. Differences in key phrases were also found during the two terms of the Bush and Obama presidencies, suggesting a shift in foreign policy priorities even under the same president. Although the limitations of this paper's methodology preclude finding any conclusive shift in foreign policy priorities by president, this paper demonstrates the feasibility of applying basic text-mining techniques in answering social science questions where data can be found in text-based sources.*

### 1. Introduction

Inspired by Cogburn and Wozniak's investigation of public remarks of U.S. Secretaries of State (2013) [1] and the author's experience as a political transcriptionist, this study is interested in establishing how and to what extent U.S. foreign policy priorities change by president by looking at transcripts of foreign policy budget hearings. The annual budgets proposed by the Executive for agencies involved in foreign affairs – the State Department, the U.S. Agency for International Development (USAID), the Peace Corps, and others – must pass relevant committees in Congress, and face scrutiny by lawmakers who approve agency funding. Transcripts, by their verbatim nature, provide rich sources of information on these hearings, and by extension on the U.S.' foreign policy priorities.

### 2. Purpose

The purpose of this project is to characterize and compare trends in foreign policy priorities by text-

mining transcripts of the House Appropriations Subcommittee on State and Foreign Operations, the congressional committee overseeing all administration spending requests related to U.S. foreign policy. With a dataset representing transcripts from 127 congressional hearings from the years 2000 to 2019, this study looks for key foreign policy phrases in the transcripts and whether their distribution differs depending on which president is in office.

Performing this study will address the feasibility of using congressional appropriations hearings to analyze foreign policy trends of administrations, mindful that other text-based sources, such as speeches, media interviews, and budget documents, can also provide information on foreign policy priorities.

And while this study is concerned with a domain-specific question in the foreign policy field, its conceptual approach – of using text-mining and automated key phrase classification to characterize the prominence of related topics and concepts – has potential applications in other fields where text-based sources are abundant.

### 3. Literature Review

Previous studies have also employed computer-assisted content analysis in the social sciences. Aldering & Vliegthart (2016) examined almost 150,000 newspaper articles with automated software to discern and measure political leaders against a set of pre-defined political leadership images, finding that computerized content analysis provided a valid measurement for leadership compared to manual analysis [2]. Albaugh et al. (2013) found it was possible to use dictionary-based content analysis to reliably measure policy agendas [3]; Kirilenko et al. (2012) used computer-assisted key word frequency analysis to show differences in public attitudes towards scientific principles among British and American newspapers [4]. And König & Finke (2015) examined dozens of German legislature bills pre- and post-9/11 to see how often political parties initiated counterterrorism legislation [5]. Although this present study takes a more

rudimentary text-mining approach to its source data than the above-mentioned studies, like those studies, it assumes a functional and expedient role for computer-assisted content analysis, capitalizing on the richness of text-based data, found in print and online, to answer political science questions in ways not possible before the advent of computer-assisted analysis.

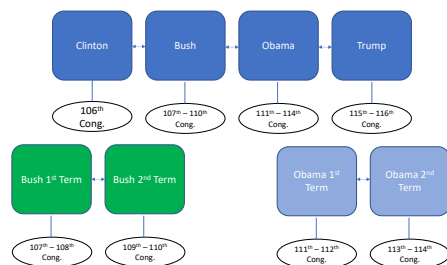
## 4. Conceptual Framework

Kirilenko et al. sought to extract the most important topics of public discourse from their dataset, by using a method combining key word selection, text modification based on a key word dictionary, and frequency and proximity analysis of the text population. In this study, I also sought the most important topics of the transcripts, as measured by frequency and key word selection, and by discounting words associated with the transcript production process such as acronyms and proper nouns referring to individual speakers.

The size of this dataset was informed by König & Finke, who analyzed 90 counterterrorism bills introduced over a 10-year period; in this study, I analyze 127 congressional hearings over a 20-year period, relatively similar in scope.

Whereas König and Finke constructed a dictionary with a large set of pre-defined terms related to their topic of interest, drawing from legal documents and thesauri, this study takes a less prescriptive approach, using the metric of term frequency-inverse document frequency of key words and phrases as the main determinant of a topic's importance. While not as refined, this approach is suitable as a starting point for characterizing the efficacy of text mining the source data.

In this study, I aggregated the hearings' transcripts by the four most recent presidents to see how foreign policy priorities change across presidents. I also aggregated transcripts of the first and second terms of Presidents Bush and Obama in order to explore changes in priorities over each president's tenure. Figure 1 demonstrates these groupings:



**Figure 1. Conceptual framework**

The agencies represented in this dataset include well-known entities like the State Department, USAID, and the Peace Corps. Other entities and programs under the committee's jurisdiction include the Export-Import Bank, the Asia Foundation, and various joint commissions on topics like international religious freedom and fisheries management. Together, these agencies and programs represent the executive's instruments of non-military foreign policy in domains such as diplomacy, international development, trade, environmental protection, law enforcement, and promotion of human rights and democracy.

## 5. Research Questions

The following questions are asked of this dataset: What is the distribution of top key phrases? How do presidential terms compare in terms of this distribution? And is there evidence of a shift in key phrases during a two-term presidency?

## 6. Methodology

The dataset consists of 127 transcripts of congressional hearings of the House Appropriations Subcommittee on State and Foreign Operations, obtained through Congressional Quarterly (CQ) Roll Call, a subscription-based provider of congressional data. Access to CQ was provided by American University (AU). AU's library website ([www.american.edu/library](http://www.american.edu/library)) links to the CQ Roll Call database, which furnishes verbatim transcripts searchable by congressional committee. The selected committee "HAPP-State-Foreign Operations" yields 128 transcripts (127 after discarding an errant Senate committee transcript) dating back to the 106th Congress (February 2000) and ending, as of this writing, with transcripts from the 116th Congress (March 2019).

**Table 1. Transcript metadata**

Congress	Year	President	No. Hearings	No. Pages	File Size (mb)
106	2000	Clinton	4	216	0.516
107	2001 - 2002	Bush	9	424	0.978
108	2003 - 2004	Bush	15	734	1.7
109	2005 - 2006	Bush	15	729	1.6
110	2007 - 2008	Bush	18	728	1.6
111	2009 - 2010	Obama	19	583	1.4
112	2011 - 2012	Obama	14	434	1
113	2013 - 2014	Obama	10	332	0.836
114	2015 - 2016	Obama	12	365	0.887
115	2017 - 2018	Trump	7	253	0.605
116	2019 - 2020	Trump	4	179	0.455
<b>Total</b>			<b>127</b>	<b>4977</b>	<b>11.577</b>

The transcripts were downloaded and exported as PDFs, resulting in 127 files as shown in Table 1.

## 6.1. Preparing the corpus

A virtual corpus of all 127 transcripts was created using the software package RStudio and the package ‘tm’. Because the transcripts are PDFs, the ‘pdftools’ package was also loaded into RStudio. The corpus then underwent transformations to remove whitespace in case key phrases had blank spaces in different positions; remove numbers (numbers in these budget hearings appear frequently, and are not considered useful for this study); remove punctuation; and to make all text lowercase. Lastly, English stopwords were removed from the corpus, as well as frequently occurring first and last names, identifiers (such as “Mr.” or “Ms.”), and abbreviations such as callouts of a representative’s home state (R-TX, for example). Other words and phrases were also removed if they appeared frequently in the corpus and were related to the transcription process (such as a copyright disclaimer inserted into each transcript by the transcription firm).

Then, a corpus was made for each president consisting of hearings occurring during their terms, with the same transformations as above, resulting in four ‘presidential’ corpuses. Finally, a corpus was made for each term of the Bush and Obama presidencies, resulting again in four additional corpuses.

## 6.2. Finding statistically significant key phrases

First, the most statistically significant key phrases are sought for the entire corpus of 127 transcripts. A document-term matrix (DTM) was created using term frequency-inverse document frequency weighting for more relevant results, and key phrases were found and sorted by statistical significance using the ‘RWeka’ package in RStudio. The results were displayed as histograms using the ‘ggplot2’ package.

Similar techniques were applied to each presidential corpus, as well as to the corpuses of each term for presidents Bush and Obama. This enabled comparison of key phrase frequency by president and by the first and second terms for Bush and Obama.

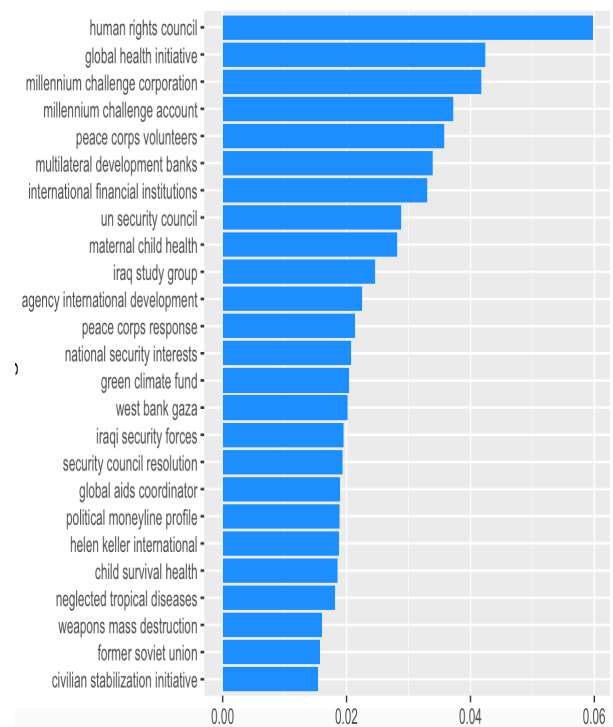
## 6.3. Finding word clusters

To find out the efficacy of using unsupervised machine learning to automatically define categories of key words, a cluster analysis of the entire corpus was performed with a *K*-means clustering algorithm. The packages ‘skmeans’, ‘clue’, ‘cluster’, and ‘fpc’ were

used, and the results were displayed in a word cloud for visual impact.

## 7. Findings

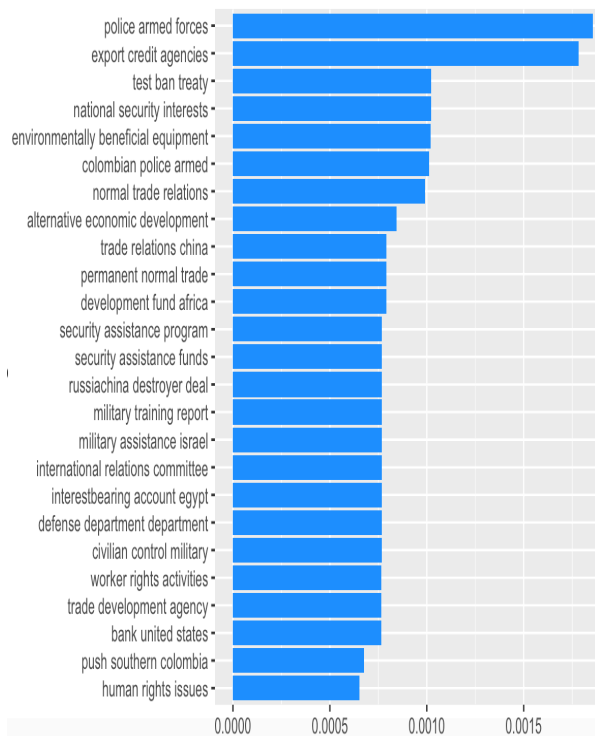
The first research question sought the distribution of key phrases over the entire corpus of four presidents. The key phrases sorted by statistical significance appear in Figure 2:



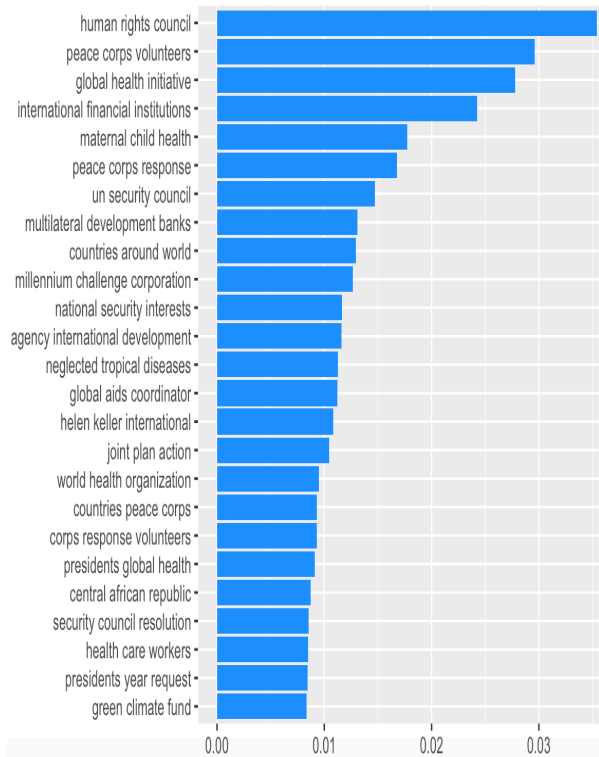
**Figure 2. Most significant phrases, 2000-2019**

The most statistically significant key phrases throughout the study period show a focus on initiatives, issues, and places like the UN Human Rights Council; the U.S.’ Global Health Initiative, an Obama-era approach to global health policy; the Peace Corps; the West Bank and Gaza; global AIDs coordination; and tropical diseases. This shows a humanitarian and human development-centric foreign policy throughout the study period.

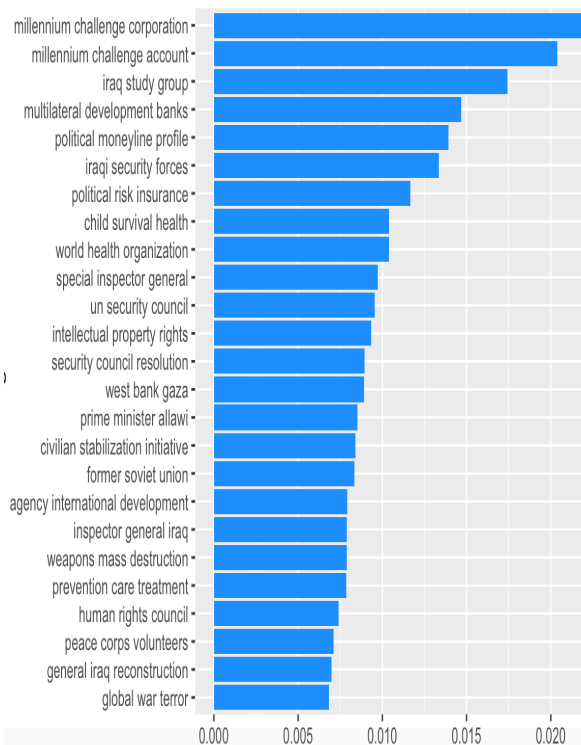
Following are figures representing the most significant key phrases by presidential administration, in response to the second research question:



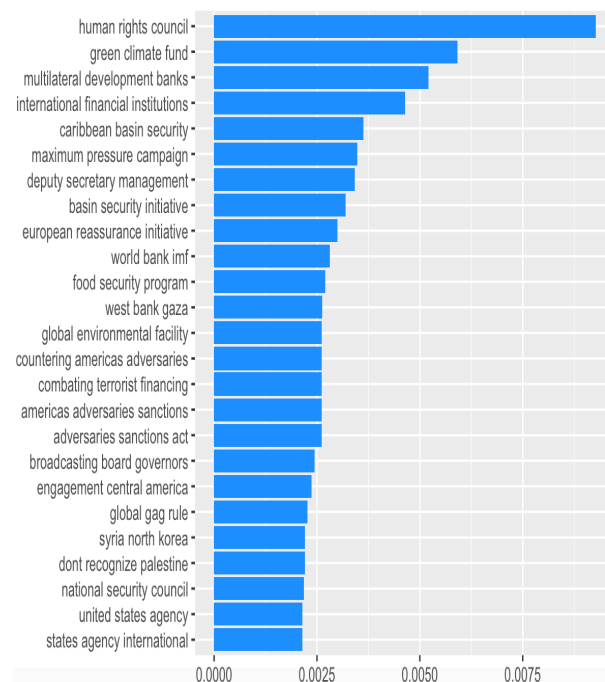
**Figure 3. Most significant phrases, Clinton administration (2000)**



**Figure 4. Most significant phrases, Obama administration (2009 – 2016)**

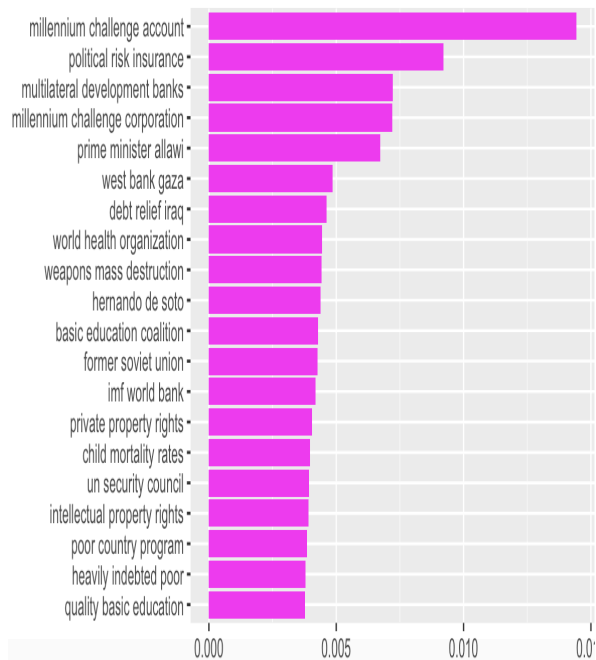


**Figure 5. Most significant phrases, Bush administration (2001 – 2008)**

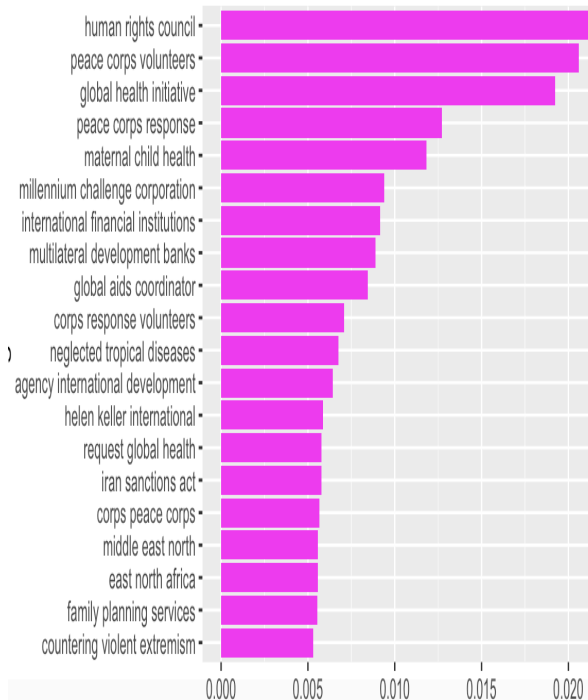


**Figure 6. Most significant phrases, Trump administration (2017 – 2019)**

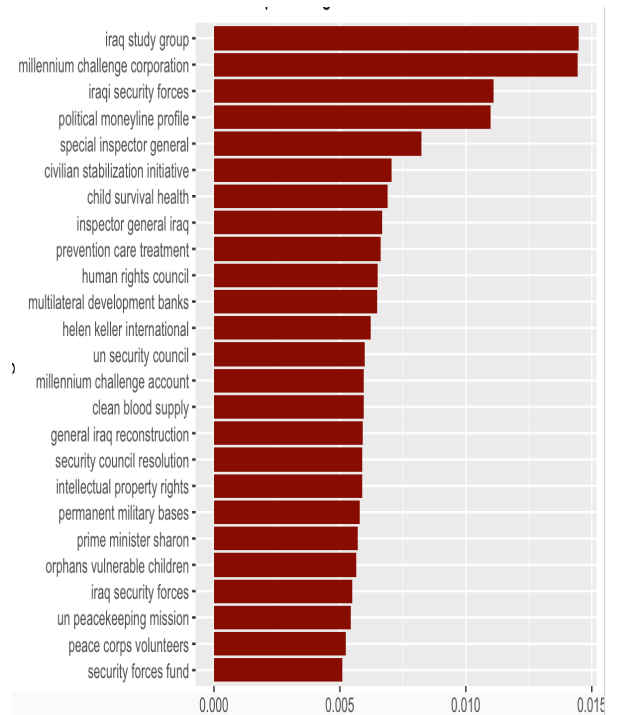
The third part of the research question focuses on shifts in foreign policy priorities from the first and second terms of a presidency. Data is available from both terms of the Bush and Obama administrations, and the results of key phrase analyses follow.



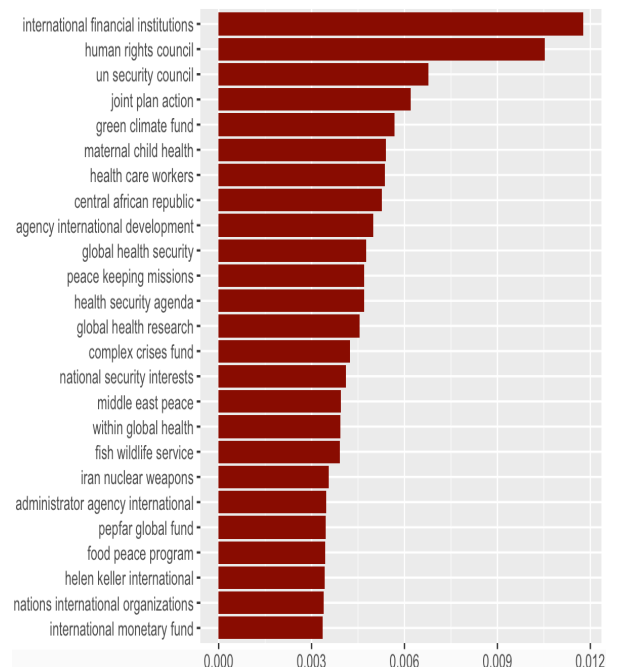
**Figure 7. Most significant phrases, Bush first term (2001 – 2004)**



**Figure 9. Most significant phrases, Obama first term (2009 – 2012)**



**Figure 8. Most significant phrases, Bush second term (2005 – 2008)**



**Figure 10. Most significant phrases, Obama second term (2013 – 2016)**



[illegible]

## 8. Discussion

During the end of the Clinton administration, the subcommittee hearings were focused on the Western Hemisphere, with key phrases associated with topics such as security assistance, drug interdiction from Latin America, international military education and training (IMET), Colombia, Bolivia, foreign military financing (FMF), and export credit agencies being the most statistically significant.

The most statistically significant key phrases during hearings in the Obama administration are related to human development, notably maternal and child health; Afghanistan and Iran; the emergence of the Joint Plan of Action, the pact between Iran and six world powers on freezing Iran's nuclear program; sanctions; and wildlife (protection).

Together, these key phrases show a divergence in topic areas, from the late Clinton-era focus on Latin America, to the Middle East and Afghanistan during the Bush administration, to foreign aid and development in the Obama administration, to a focus on security challenges around the world in the early Trump period. It should be noted, however, that this study's methodology accounts for the frequency of key terms and does not provide context on who said them – for instance, frequent mentions of a topic during a hearing may originate from lawmakers' questions rather than from actual administration priorities.

Shifts in foreign policy priorities become apparent even during the same president's tenure. During President Bush's tenure, there is a change in key phrase significance from his first and second terms, from global security topics towards terms related to reconstruction and the U.S. presence in Iraq. The Millennium Challenge Corporation persists as a significant key phrase in both terms, though; multilateral development banks, intellectual property rights, and the UN Security Council also persist across terms, while the former Soviet Union as a key phrase drops in significance in the second term.

### 8.3 Automatic Clustering of Keywords

Page 750

nations, while also emphasizing sanctions against Iran and strategic competition with countries like China and Russia. This result shows the potential of automatic clustering of key phrases within a data set to help researchers answer questions about large data sets, not just in the foreign policy realm but in any field.

## 9. Limitations

This study is constrained by the reliance on statistical significance of key phrases alone as a measurement of foreign policy priority. Although the results illuminate the most relevant issues of a hearing, they do not indicate whether these topics translated into actual foreign policy. Moreover, the results do not show where the top key phrases originate from – the agencies or the lawmakers who approve the agencies' funding requests.

While this study has established the potential of text-mining transcripts for discerning foreign policy priorities, it could further benefit from the approaches used by previously cited authors, such as proximity and factor analysis, and cross-checking results against human-coded data.

And as Table 1 shows, only four transcripts are available for the Clinton administration. The lack of transcripts from before this period, as well as the proprietary nature of the transcripts themselves (requiring commercial or institutional access), limit the data set's effectiveness as a source for computer-assisted content analysis.

## 10. Future Research

With more historical data from past committee hearings, more light can be shed on how the government responds to contemporaneous foreign policy challenges and issues, and whether other factors related to the person of the president, such as political party affiliation, are determinants of foreign policy priorities.

Other studies could determine the effect of the congressional committee's makeup on foreign policy priorities, such as how its political party composition or average length of tenure in Congress shape foreign policy priorities.

Text mining can also be performed on other modes of foreign policy transmission, such as presidential speeches, interviews, and press briefings. These analyses can be compared with studies like this one to determine differences in key phrase frequencies by medium, to see if, for example, stated foreign policy priorities change depending on the medium and intended audience.

## 11. References

- [1] D. Cogburn and A. Wozniak, "Computationally Intensive Content Analysis of Public Diplomacy Data: Understanding the Public Remarks of US Secretaries of State, 1997-2011", 46<sup>th</sup> Hawaii International Conference on System Sciences, 2013.
- [2] L. Aaldering and R. Vliegthart, "Political leaders and the media: Can we measure political leadership images in newspapers using computer-assisted content analysis?", *Quality & Quantity*, Vol. 50, no. 5, pp. 1871-1905.
- [3] Q. Albaugh, J. Sevenans, S. Soroka, and P.J. Loewen, "The automated coding of policy agendas: A dictionary-based approach", 6<sup>th</sup> Annual Comparative Agendas Conference, Antwerp, Belgium, June 2013.
- [4] A. Kirilenko, S. Stepchenkova, R. Romsdahl, and K. Mattis, "Computer-assisted analysis of public discourse: A case study of the precautionary principle in the US and UK press", *Quality & Quantity*, Vol. 46, no. 2, pp. 501-522.
- [5] T. König and D. Finke, "Legislative governance in times of international terrorism", *Journal of Conflict Resolution*, Vol. 59, no. 2, pp. 262-282.